

УДК 81'374

СПОСОБЫ СОЗДАНИЯ ТОНАЛЬНЫХ СЛОВАРЕЙ

Д. С. СЕМАК

Гомельский государственный технический университет имени П. О. Сухого
Гомель, Беларусь

Введение. Активное развитие в настоящее время социальных сетей, блогов и форумов привело к увеличению интереса, как со стороны научного сообщества, так и со стороны многих организаций к задаче автоматического анализа мнений пользователей Интернета по различным вопросам (отношение к товарам, услугам, событиям, высказываниям). Одной из основных проблем при анализе мнений является классификация текстов по тональности. Необходимое условие для анализа тональности – составление словарного списка тональной лексики.

В данной работе рассматриваются способы создания тональных словарей.

Основная часть. Анализ тональности является одним из направлений области автоматической обработки текста на естественном языке. Под анализом тональности понимается определение отношения автора к некоему объекту действительности [1, с. 519]. Это отношение определяется по бинарной шкале («позитивное-негативное») или монополосной шкале (например, пятибалльная шкала оценки). Системы автоматического анализа тональности могут быть однодоменными (то есть относящиеся к одной предметной области) или мультидоменные (не относящиеся к определенной предметной области). Определение оценочного отношения может быть объектно-ориентированным и аспектно-ориентированным. Под первым понимается анализ тональности некоего объекта в целом, под вторым – набора характеристик данного объекта

Тип системы и постановка задачи определяет метод извлечения оценочной информации. Сегодня существует множество подходов к анализу тональности. К ним можно отнести машинное обучение с учителем и без, правила и тональные словари. Данные методы не являются взаимоисключающими. Каждый из них, обладая рядом достоинств и недостатков, может применяться совместно с другим методом.

Особое внимание заслуживает метод анализа оценочной информации с применением тонального словаря. Суть метода заключается в создании специального словаря оценочной лексики, характерной для конкретного домена. Как правило, словарь состоит из слов, словосочетаний и устойчивых выражений с указанной тональностью (положительной или отрицательной). Автоматическая система использует информацию из тонального словаря для анализа. К недостаткам данного метода можно отнести сравнительно длительное время создания. С другой стороны, качественно скомпилированный тональный словарь показывает высокую точность анализа.

Существует множество методов создания списка оценочных слов, но концептуальная модель содержит три этапа:

- 1) наполнение словаря оценочной лексикой;
- 2) приписывание лексическим единицам тональности и, иногда, условий ее реализации;
- 3) приписывание валентности оценочным словам (опционально).

Для выполнения первого этапа может применяться словарь синонимов. Очевидно, что слова-синонимы будут иметь одинаковую тональность. Но, составленный таким образом словарь не будет отражать лингвистическую специфику конкретной предметной области. Например, слово *scary* («страшный») будет иметь отрицательную тональность, но будучи употребленным в отзыве на фильм ужасов тональность будет положительной. Поэтому, для выявления подобных случаев применяется метод анализа корпуса текстов по конкретной области действительности. Извлеченная из корпуса оценочная лексика будет отражать специфику домена.

Для реализации второго этапа используется информация из словаря синонимов или контекстуальная обусловленность употребления лексики в текстовом корпусе. Лингвистическая специфика каждой предметной области диктует необходимость записывать в некоторые словарные статьи различные правила или шаблоны. Например, в текстовом корпусе оценочных отзывов на предприятия гостиничного типа слово *old* («старый») имеет негативную тональность только в том случае если оно относится к конкретным аспектам отеля (*old room, old bed ...*). Таким образом, в предложении *The hotel is located in an old part of town* («Отель находится в старой части города») слово *old* не имеет тональности. Эта специфика отражена в словарной статье в виде правила, которое называется *условие реализации тональности*.

Выполнение третьего этапа является опциональным и требуется только в случае обеспечения тех автоматических систем, которые дополнительно определяют на сколько положительным или отрицательным является высказывание

Заключение. Исходя из вышеизложенных данных, можно сделать вывод о том, что для создания качественного тонального словаря требуется применить комплексный подход.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. **Kushal, D.** Mining the peanut gallery: opinion extraction and semantic classification of product reviews / D. Kushal, L. Steve, M. David // Proceedings of WWW, 2003. – P. 519–528.